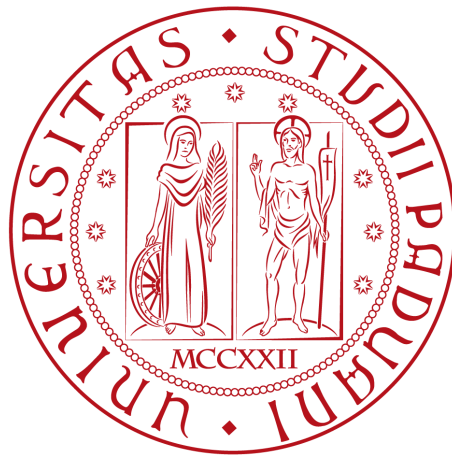


**Università degli Studi di Padova**

Dipartimento di Fisica ed Astronomia “Galileo Galilei”  
Corso di Laurea in Fisica

Tesi di Laurea Triennale



# Crescita di colonie batteriche bioluminescenti: ottimizzazione di modelli teorici

**Relatore:**  
**Prof. Antonio Trovato**

**Laureando:**  
**Francesco Fanchin**

**Anno Accademico 2015/2016**

# Indice

1	Introduzione	3
2	Il modello matematico della bioluminescenza	6
3	Descrizione generale del problema	9
4	Il metodo di Nelder-Mead	11
5	Il simulated annealing	13
6	Aspetti tecnici del programma e risultati	16
7	Conclusioni	20
	Bibliografia	21

# Abstract

Cellule batteriche di una singola colonia comunicano fra di loro secondo il meccanismo del Quorum Sensing, riuscendo a coordinare comportamenti collettivi quali ad esempio la bioluminescenza.

Nel tipico meccanismo di Quorum Sensing, molecole segnale vengono prodotte dai batteri; al di sopra di una concentrazione di soglia una frazione consistente di molecole segnale si lega a recettori presenti nei batteri, attivando quindi un nuovo fenotipo, per esempio la bioluminescenza, tramite una cascata di segnali cellulari.

In questo lavoro di tesi, si propone di provare a riprodurre misure sperimentali di bioluminescenza, ottenute da un gruppo dell'Università del Salento, per mezzo di un modello teorico semplificato, che tiene conto della crescita dei batteri, della produzione e degradazione delle molecole segnale, e della soglia e cooperatività del meccanismo di attivazione del Quorum.

I diversi parametri presenti nel modello verranno ottimizzati minimizzando la differenza fra previsione teorica e dati sperimentali per mezzo del metodo del "simulated annealing".

# Capitolo 1

## Introduzione

Il *quorum sensing* (**QS**) è un comportamento collettivo coordinato presente in numerosi organismi viventi, in particolare nei batteri: la sua azione deriva da fattori fisico-geometrici come la densità cellulare, il clustering (raggruppamento in microcolonie) e il "trasferimento di massa", ovvero la diffusione di fluidi extracellulari [1].

Dal punto di vista biochimico, questi batteri sintetizzano e secernono nell'ambiente esterno specifiche molecole segnale denominate *autoinduttori* (**AI**): esse permettono a una colonia di trarre informazioni sulla propria densità spaziale. Nel caso in cui la crescita batterica superi una certa soglia, denominata quorum, gli **AI**, legandosi ai recettori presenti nei batteri, alterano l'espressione genica innescando determinati processi cellulari dipendenti dai geni coinvolti.

Il **QS** è inoltre responsabile della degradazione delle molecole segnale al raggiungimento di limiti ambientali come la disponibilità di nutrienti o di spazio per la crescita della popolazione batterica. Sperimentalmente sono stati osservati fenomeni di feedback positivo e negativo<sup>1</sup> a varie concentrazioni di AI: nel caso di *Vibrio Harvey* (oggetto del seguente esperimento) è prevalente il primo.

Un esempio vistoso di **QS** è rappresentato dalla *bioluminescenza*: alcune specie batteriche sono in grado di emettere luce in modo sincronizzato. La reazione chimica coinvolge un'enzima, la luciferasi, e richiede la presenza simultanea di alcuni composti organici ( $FMNH_2, RCOH$ ) e ossigeno molecolare. Fra i prodotti finali vi è l'emissione di un fotone nel range blu-verde:



La ricerca si articola da un lato nella modellizzazione matematica del processo, dall'altro nella determinazione di parametri sperimentali di inte-

---

<sup>1</sup>Il tasso di produzione o degradazione delle molecole segnale cambia con il superamento della soglia del quorum.

resse biologico: in quest'ultima direzione prende corpo un esperimento di P. Alifano e D. Delle Side [2] dell'Università del Salento volto a investigare la distribuzione del tempo di attivazione dell'emissione bioluminescente. Nello specifico, é stata prelevata una colonia batterica di *Vibrio Harvey* di dimensioni dell'ordine del mm e inserita in una camera climatica completamente oscurata a temperatura e pressione costanti; un tubo fotomoltiplicatore (PMT) montato al di sopra della colonia rivelava la quantità di luce emessa; un sistema di acquisizione dati, collegato al tubo, produceva un segnale proporzionale alla corrente elettrica generata.

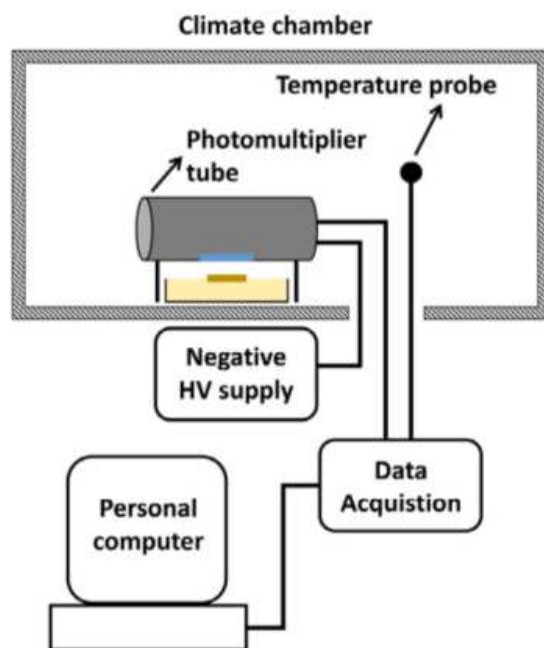


Figura 1.1: Schema dell'apparato sperimentale

La grandezza rivelata dal PMT é il *flusso radiante*, ovvero l'energia emessa sotto forma di radiazione nell'unità di tempo: moltiplicando per il tempo di acquisizione( 10 minuti) si ricava una stima accurata del numero di fotoni emessi. Secondo gli autori dell'esperimento, l'emissione bioluminescente potrebbe seguire la distribuzione di Gumbel [3], la quale interviene nello studio di eventi "estremi": esempi di applicazioni includono il livello massimo dei fiumi in un certo intervallo temporale, la magnitudo dei terremoti o la valutazione del rischio finanziario.

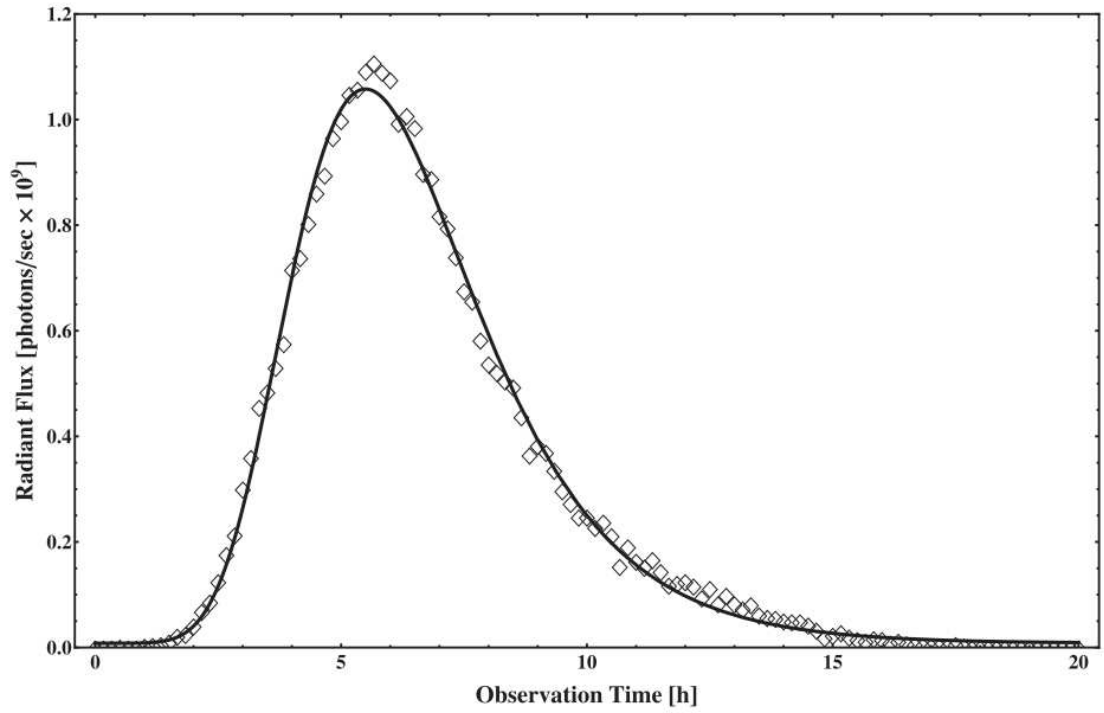


Figura 1.2: Fitting dei dati sperimentali con una distribuzione di Gumbel

Nel capitolo seguente verrà introdotto un modello matematico in cui l'emissione bioluminescente dipenderá da piú fattori concomitanti e connessi: esso sarà alla base del tentativo di riproduzione dei dati sperimentali.

## Capitolo 2

# Il modello matematico della bioluminescenza

Le equazioni differenziali sono uno strumento matematico utile a descrivere relazioni tra fenomeni variabili nel tempo: nel problema in esame sono presenti quattro grandezze dipendenti dal tempo e per semplicità si trascurano eventuali dipendenze spaziali: ciò significa che le variazioni delle grandezze perturbano uniformemente la colonia. Al tempo  $t$ , avremo dunque:

- $n(t)$ , la concentrazione di batteri;
- $c(t)$ , la concentrazione di molecole segnale;
- $p(t)$ , la concentrazione di batteri attivati dalle molecole segnale;
- $f(t)$ , l'emissione luminosa effettivamente osservata.

$n(t)$  deve essere una funzione crescente in  $t$ , per esempio una distribuzione cumulativa: nello specifico, la scelta è ricaduta sulla *funzione di Gompertz*, utilizzata in microbiologia per il logaritmo della curva di crescita batterica e sperimentalmente comprovata:

$$n(t) = n_0 10^{\kappa \exp[-\exp(1 + \frac{e\mu}{\kappa}(\lambda - t))]}$$

con  $n_0, \kappa, \lambda, \mu$  parametri incogniti.

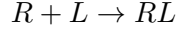
Al fine di riprodurre la fenomenologia del **QS**, si introduce una dipendenza di  $c(t)$  da  $n(t)$  tramite un'equazione differenziale lineare; inoltre si assume che le molecole segnale degradino spontaneamente nel tempo con un meccanismo analogo a quello di un decadimento: tradotto in formule,

$$\frac{dc(t)}{dt} = \alpha n(t) - \beta c(t);$$

$\alpha$  e  $\beta$  sono altri due parametri incogniti, rispettivamente il tasso di produzione e il tasso di degradazione delle molecole segnale.

Il processo di ricezione degli **AI** da parte dei batteri non é immediato, al crescere del tempo  $t$  aumenterà il numero di batteri "attivati": l'emissione bioluminescente sarà dunque legata al processo di legame degli AI ai recettori nelle cellule batteriche.

La *funzione di Hill* [4] descrive la reazione di equilibrio fisico-chimico tra due composti. Data la reazione:



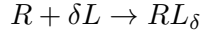
con  $R$  il recettore e  $L$  il ligando<sup>1</sup>, si vuole esprimere la frazione  $y$  di recettori legati  $RL$  sul numero totale di recettori in funzione della costante di dissociazione  $c^*$ . Occorre dunque risolvere il sistema:

$$\begin{cases} y = \frac{[RL]}{[R] + [RL]} \\ c^* = \frac{[R][L]}{[RL]} \end{cases}$$

Eliminando  $[RL]$ , si ottiene:

$$y = \frac{[L]}{[L] + c^*}$$

Nel caso di legame multiplo del tipo



che descrive per  $\delta$  intero e maggiore di uno un processo di legame cooperativo di  $\delta$  ligandi al recettore, é possibile pervenire a una formula analoga:

$$y = \frac{[L]^\delta}{c^* + [L]^\delta}$$

e, dividendo per  $[L]^\delta$ ,

$$y([L]) = \frac{1}{1 + (\frac{c^*}{[L]})^\delta}$$

Quest'ultima é nota come *funzione di Hill* e si utilizza con  $\delta$  non intero per descrivere processi di legame caratterizzati da stati intermedi.

Il parametro  $\delta$  misura la *cooperatività* del processo. Poiché al tempo  $t$  sono presenti  $n(t)$  batteri, si ipotizza che la frazione di cellule attivate  $p(t)$  sia uguale per ogni tempo  $t$  alla frazione di recettori legati (trascurando così la durata del processo cellulare di attivazione della bioluminescenza, di circa  $20 \div 30$  minuti) .Si definisce dunque la seguente funzione:

$$p(t) = \frac{n(t)}{1 + (\frac{c^*}{c(t)})^\delta}$$

---

<sup>1</sup>ovvero le molecole segnale.



con  $c(t)=[L]$  e  $n(t)$  la concentrazione di batteri al tempo  $t$ ; si noti la dipendenza crescente da  $n(t)$  e  $c(t)$  (per  $\delta > 0$ ), in accordo con la teoria. Anche in questa equazione vi sono due parametri incogniti:  $c^*$ , ovvero la soglia del quorum in termini di concentrazione di molecole segnale, e  $\delta$ . La quantità fisicamente osservata sarà la derivata di  $p(t)$ , moltiplicata per un fattore  $A$ :

$$f(t) = A \frac{dp(t)}{dt}$$

In totale vi sono nove parametri da determinare sulla base dei risultati sperimentali: con opportuni riscalamanti é possibile ridurre il problema a sette dimensioni. Ponendo:

$$\begin{cases} \tilde{\alpha} = \frac{n_0 \alpha}{c^*} \\ \tilde{A} = A n_0 \end{cases}$$

e introducendo le nuove funzioni:

$$\begin{cases} \widetilde{n(t)} = \frac{n(t)}{n_0} \\ \widetilde{c(t)} = \frac{c(t)}{c^*} \\ \widetilde{p(t)} = \frac{p(t)}{n_0} \end{cases}$$

si ottiene un sistema di equazioni analogo a quello precedente dove l'osservabile sperimentale  $f(t)$  rimane invariato. Tralasciando le  $\sim$  si ottiene quindi il modello definitivo, che dipende da sette parametri:

$$\begin{cases} n(t) = 10^{\kappa * \exp[-\exp(1 + \frac{c\mu}{\kappa}(\lambda - t))]} \\ \frac{dc(t)}{dt} = \alpha n(t) - \beta c(t); \\ p(t) = \frac{n(t)}{1 + (\frac{1}{c(t)})^\delta} \\ f(t) = A \frac{dp(t)}{dt} \end{cases}$$

Nel capitolo successivo verrà descritto l'algoritmo utilizzato per trovare soluzioni di questo sistema compatibili con le curve sperimentali.

## Capitolo 3

# Descrizione generale del problema

Il problema dell'interpolazione di dati sperimentali con una curva si può tradurre in un problema di ottimizzazione, ovvero ricerca di massimi/minimi globali, mediante il *metodo dei minimi quadrati*: di seguito se ne riassumono brevemente le caratteristiche, senza troppa pretesa di rigore matematico. Dati  $N$ ,  $m$  numeri interi, si ipotizza che una funzione dipendente da  $m$  parametri incogniti  $a_1, \dots, a_m$  e con dominio unidimensionale  $f_{(a_1, \dots, a_m)}^{th}(t)$  possa interpolare dei dati sperimentali  $f^{exp}(t_i)$ ,  $i=1, \dots, N$ ; anche  $f^{exp}$  ha dominio unidimensionale.

Si definisce una funzione opportuna  $E(a_1, \dots, a_m)$  il cui punto di minimo globale corrisponda alla migliore approssimazione della serie di dati sperimentali: come criterio di "vicinanza" si sceglie la somma dei quadrati delle differenze fra i punti sperimentali e quelli teorici.

Tradotto in formule,

$$E(a_1, \dots, a_m) = \sum_{i=1}^N [f^{exp}(t_i) - f_{(a_1, \dots, a_m)}^{th}(t_i)]^2$$

è la funzione da minimizzare.

Nel caso in cui si riesca a trovare il punto di minimo globale, si determinano i parametri  $a_1, \dots, a_m$  i quali a loro volta definiscono univocamente la funzione di best-fit.

Procedendo per via analitica, ovvero imponendo le derivate parziali nulle, la dimensione del dominio e la struttura matematica della funzione (lineare/non lineare) incidono notevolmente sulle possibilità di successo del metodo: in particolare, alcuni problemi non sono risolvibili attualmente in questo modo e si cercano soluzioni approssimate con l'ausilio del computer.

All' aumentare della complessità del problema, può essere ragionevole passare dalla ricerca della migliore soluzione alla ricerca di una serie di buone

soluzioni ritenute sufficientemente "vicine" al minimo. Fatta questa premessa, occorre capire quale algoritmo utilizzare per ottimizzare il nostro modello descritto nel capitolo precedente: vi sono sette parametri incogniti e la funzione  $f(t)$ , a meno di soluzioni banali, é evidentemente non-lineare.

Un'ulteriore complicazione risiede nel calcolo di  $f(t)$  in cui sono presenti due equazioni differenziali: piú avanti saranno descritti i metodi per risolverle. In generale ci si può aspettare che  $E$  non sia affatto monotona e che, data la vastità dello spazio dei parametri, c'è il rischio che i metodi deterministici "tradizionali", basati cioè sull' accettazione reiterata di punti col minor valore della funzione, si fermino a un minimo locale troppo distante da una buona soluzione: inoltre, algoritmi di questo tipo potrebbero essere fortemente dipendenti dal punto di partenza iniziale.

In ultima analisi, sono necessari dei metodi in grado di accettare anche punti con un valore della funzione maggiore: ciò può sembrare controproducente, in realtà questo consente di esplorare maggiormente lo spazio dei parametri, permettendo di trovare, almeno in linea di principio, punti di minimo caratterizzati da un minor bacino di attrazione; la difficoltà sta nel garantire la convergenza e la stabilità del metodo, almeno per un numero di iterazioni sufficientemente lungo. L'algoritmo utilizzato consiste nell'unione di due algoritmi: il metodo di Nelder-Mead e il "simulated annealing", rispettivamente uno deterministico e l'altro stocastico.

## Capitolo 4

# Il metodo di Nelder-Mead

In questo capitolo verrà descritto il metodo deterministico che sta alla base della ricerca del minimo globale: successivamente saranno aggiunti altri miglioramenti. Come nel capitolo precedente, sia  $E(a_1, \dots, a_m)$  la funzione da minimizzare.

Il *metodo di Nelder-Mead* [5] si basa su un insieme di  $m+1$  punti che possono variare la loro mutua distanza secondo regole precise: nelle vicinanze di un minimo i punti tendono a raggrupparsi sempre di più attorno a esso all'aumentare delle iterazioni, fino a "coincidere" (in pratica, dal punto di vista computazionale, serve una condizione di terminazione dell'algoritmo: si può imporre sulla distanza fra i punti o sui valori delle funzioni calcolate in quei punti). L'insieme dei punti è detto *simplex*, mentre l'algoritmo viene anche soprannominato "amoeba", immaginando un ipotetico organismo proteiforme, formato dai punti del *simplex*, che esplora lo spazio dei parametri.

Dato un *simplex* iniziale, il metodo consiste in un ciclo composto da una serie di steps:

- Si calcola il valore della funzione da minimizzare per ogni punto del *simplex*;
- Si identificano il punto migliore (cioè con il valore minimo della funzione), il punto peggiore e il secondo punto peggiore, rispettivamente  $x_1$ ,  $x_m$ ,  $x_{m+1}$ ;
- Si calcola il baricentro  $x_0$  di tutti i punti escluso il peggiore;
- Si trova il punto "riflesso", dato dalla formula

$$x_R = x_0 + \alpha(x_0 - x_{m+1})$$

Se  $E(x_1) < E(x_R) < E(x_m)$  si sostituisce  $x_{m+1}$  con  $x_R$  e si torna al punto 1;

- Se  $E(x_R) < E(x_1)$  si determina il punto "espanso"

$$x_E = x_0 + \gamma(x_0 - x_{m+1})$$

Se  $E(x_E) < E(x_R)$ , ovvero se il punto espanso é migliore di quello riflesso, si sostituisce  $x_{m+1}$  con  $x_E$  e si torna al punto 1. In caso contrario si sostituisce  $x_{m+1}$  con  $x_R$  e si torna anche in questo caso al punto 1.

- L'unico caso rimasto é quello di  $E(x_R) \geq E(x_m)$ : si calcola il punto "contratto"

$$x_C = x_0 + \rho(x_0 - x_{m+1})$$

Se il punto contratto é migliore del punto peggiore, cioè se  $E(x_C) < E(x_{m+1})$ , si sostituisce  $x_{m+1}$  con  $x_C$  e si torna al punto 1, altrimenti si va al punto successivo

- Tutti i punti tranne  $x_1$  vengono sostituiti tramite la formula:

$$x_i = x_1 + \sigma(x_i - x_1)$$

con  $i=2, \dots, m+1$ , e il ciclo ricomincia.

I coefficienti  $\alpha$ ,  $\gamma$ ,  $\rho$ ,  $\sigma$  sono in linea di principio arbitrari: valori tipici, adottati in questa tesi, sono  $\alpha=1$ ,  $\gamma=2$ ,  $\rho=0.5$ ,  $\sigma=-0.5$ . Questo metodo da solo é troppo dipendente dal simplesso iniziale, inoltre le soluzioni trovate sono decisamente migliorabili. Per ovviare a questi inconvenienti si aggiunge il contributo stocastico del "simulated annealing".

## Capitolo 5

# Il simulated annealing

Il seguente algoritmo [6] si basa su un procedimento adottato in scienza dei materiali, la ricottura (annealing in inglese).

Un solido, in generale, può non essere composto da un reticolo cristallino e presentare una struttura disordinata (solido amorfo o vetroso).

Un modo per cristallizzarlo consiste nel portarlo a una temperatura molto alta e abbassarla lentamente. La spiegazione deriva dalla meccanica statistica: assumendo valida in prima approssimazione la *statistica di Maxwell-Boltzmann*, la probabilità di ogni possibile configurazione  $\{r_i\}$  è pesata dal *fattore di Boltzmann*  $e^{-\frac{E(r_i)}{KT}}$ , con  $E(r_i)$  l'energia del sistema e  $K$  la *costante di Boltzmann*.

Dalla formula si può dedurre che al diminuire di  $T$  il sistema tende alle configurazioni di energia minima. Questa però non è una condizione sufficiente: a seconda della struttura di  $E(r_i)$ , se la temperatura scende troppo velocemente, il sistema può rimanere "intrappolato" in un minimo locale di energia, che in linea di principio può essere distante dal minimo globale; viceversa, un raffreddamento lento permette di esplorare una maggior varietà di configurazioni e di giungere a valori più bassi.

Nella seguente figura è mostrato un esempio di curva di energia monodimensionale:

Osservando la posizione del punto rosso, che rappresenta il sistema fisico, un raffreddamento repentino aumenterebbe il *fattore di Boltzmann* delle configurazioni del minimo locale più vicino; una volta raggiunto il minimo locale, le configurazioni adiacenti sarebbero caratterizzate da un fattore di Boltzmann molto più basso, dunque da una minor probabilità di "risalita". Nel caso contrario, il *fattore di Boltzmann* non varia troppo e il punto rosso ha una probabilità maggiore di accedere ad altre configurazioni, sperabilmente più favorevoli in energia.

Il metodo del *simulated annealing* è legato indissolubilmente all'*algoritmo di Metropolis*, ideato per generare numeri casuali che si distribuiscano secondo una certa distribuzione nota a priori; qualora si utilizzi la *distribuzione di*

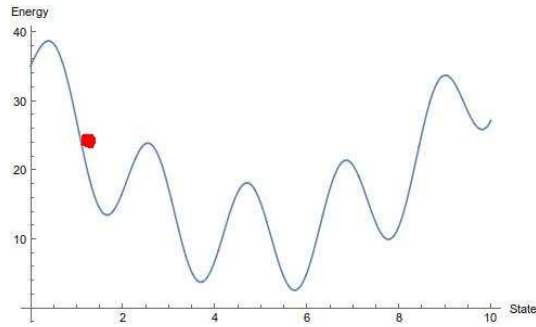


Figura 5.1: Esempio di sistema in una curva di energia monodimensionale

*Maxwell-Boltzmann*, quest'ultimo mantiene la temperatura costante.

La novità del **SA** risiede proprio nella diminuzione opportuna del parametro  $T$ . Di seguito la spiegazione dell'algoritmo:

- Si sceglie una configurazione  $\{r_i^A\}$  e si calcola  $E_A = E(r_i^A)$ ;
- Si cambia configurazione, per esempio  $\{r_i^B\}$  e si calcola  $E_B = E(r_i^B)$ ;
- Si determina il rapporto tra i fattori di Boltzmann di  $\{r_i^B\}$  e  $\{r_i^A\}$ , ovvero

$$P(\Delta E) = e^{-\frac{(E_B - E_A)}{KT}}$$

Qui si aprono due sottocasi:

- se  $P(\Delta E) > 1$ ,  $\{r_i^B\}$  viene accettata come nuova configurazione e si torna al punto 1;
- in caso contrario si genera un numero pseudocasuale  $a$  compreso tra 0 e 1; se  $P(\Delta E) > a$ ,  $\{r_i^B\}$  viene accettata, altrimenti si lascia  $\{r_i^A\}$  come configurazione. In entrambi i casi si torna al punto 1.

Il ciclo va avanti in questo modo fino a un certo numero di mosse  $m$ : a questo punto si abbassa il parametro  $T$ . Sono possibili varie soluzioni: in quella "geometrica", qui utilizzata,  $T$  si abbassa di un fattore  $C_R$  costante (cooling rate). Sia  $T_n$  la temperatura  $n$ -esima; denotando con  $\lfloor \dots \rfloor$  la parte intera e  $T_0$  la temperatura iniziale, si ha la seguente formula per  $T_n$ :

$$T_n = C_R^{\lfloor \frac{n}{m} \rfloor} T_0$$

Una volta nota la formula di abbassamento della temperatura, le costanti che definiscono univocamente l'annealing (in questo caso,  $m$ ,  $C_R$  e  $T_0$ ) costituiscono quella che in letteratura viene chiamata *annealing schedule*.

Rimane solo da combinare i due metodi finora descritti: mantenendo la "struttura-base" dell' algoritmo di Nelder-Mead, ogni qualvolta si calcoli la funzione costo  $E$  per un vertice del semplice, si aggiunge la quantità  $-T \log(b)$ , con  $b$  un numero casuale distribuito uniformemente,  $0 < b < 1$ ; se invece si tratta di una "mossa" (riflessione, espansione, ...), la stessa quantità va sottratta (il numero casuale deve cambiare). Ogni  $m$  mosse  $T$  diminuirà secondo la particolare *annealing schedule* utilizzata.

In pratica, si è introdotto un rumore statistico distribuito secondo la Maxwell-Boltzmann e proporzionale alla temperatura: ci sarà una probabilità non nulla che l'algoritmo accetti mosse energeticamente sfavorevoli.

La convergenza è comunque garantita dal segno dei due contributi e dall'abbassarsi della temperatura; tuttavia persiste il problema pratico della dipendenza del tempo di convergenza dall'*annealing schedule* e dalle condizioni iniziali.



## Capitolo 6

# Aspetti tecnici del programma e risultati

Nella descrizione generale del problema é emersa la necessità di trovare una serie di buone soluzioni ritenute sufficientemente vicine al minimo globale della funzione-costo  $E$ : in concreto, questo ha implicato la realizzazione di un programma in C che implementasse l'algoritmo sopramenzionato.

Il semplice iniziale é rappresentato da una matrice  $8 \times 7$ ; per evitare instabilità dell'algoritmo, i punti del semplice sono raggruppati attorno a un punto centrale con un contributo pseudocasuale.

La distanza fra i vari punti dipende strettamente dalla natura del problema e sono necessarie alcune prove per trovare il range corretto.

Altri parametri da impostare sono quelli relativi all'*annealing schedule*: dopo alcuni tentativi, la scelta é ricaduta su  $T_0=2$ ,  $m=10$  e  $cool=0.99$ . Con  $T_0$  troppo alti si rischia di "perturbare" eccessivamente amoeba, che può anche divergere a causa della struttura di  $E$ ; con  $T_0$  troppo bassi vengono accettate poche "risalite" e si ricade nel problema degli algoritmi deterministici. Il cooling rate é il parametro più importante che garantisce la "capillarità" nella ricerca di minimi.

Empiricamente, si é visto che oltre un certo valore di  $m$  non si ottengono miglioramenti apprezzabili perché amoeba continua la ricerca sempre nella stessa porzione di spazio dei parametri: vale il viceversa per un' alta cooling rate (anche  $cool = 0.9999$ , ma il tempo di calcolo aumenta a dismisura).

Il calcolo della funzione-costo é affidato a una funzione apposita, separata dal resto del programma: in essa é presente un array di punti sperimentali. Per discretizzare  $c(t)$ , é stato utilizzato l'algoritmo di Eulero, mentre per  $f(t)$  é stata introdotta una derivata centrata. In formule,

$$c[i + 1] = c[i] - dt * (\beta * c[i] - \alpha * n[i]);$$

$$f[i] = A * (0.5 * (p[i + 1] - p[i - 1])/dt);$$

La scelta dell'unità di tempo  $dt$  per calcolare le varie funzioni é delicata: da un lato, al diminuire di  $dt$  il tempo di calcolo aumenta, dall'altro l'errore commesso dagli algoritmi a ogni passo é proporzionale a  $dt^2$  e questo può generare instabilità; prestando attenzione al semplice iniziale, é stato impostato  $dt=600$  s per giungere a una serie di risultati in un tempo ragionevole. L' algoritmo termina quando la differenza fra il massimo e il minimo valore di  $E$  fra i punti del semplice é minore di un certa quantità epsilon: per garantire un buon confronto visivo dei grafici senza allungare troppo i tempi di calcolo si é impostato  $\epsilon = 0.001$ .

Una volta trovato il semplice finale, se ne calcola il baricentro; questo vettore, unito al corrispondente valore in energia, viene memorizzato nella riga di un'altra matrice. Il ciclo ricomincia con un nuovo semplice iniziale generato da un punto centrale randomizzato, per ottenere una certa variabilità di soluzioni.

La matrice contenente tutti i risultati viene passata in input a una funzione che calcola, per ogni soluzione,  $E_{PDF}$ , coincidente con  $E(a_1, \dots, a_m)$  della formula dei minimi quadrati, la funzione-costo della distribuzione cumulativa (i punti  $p^{exp}(t_i)$  vanno ricavati da  $f^{exp}(t_i)$ )

$$E_{CDF}(a_1, \dots, a_m) = \sum_{i=1}^N [p^{exp}(t_i) - p_{(a_1, \dots, a_m)}^{th}(t_i)]^2$$

e infine il chi-quadrato di  $f(t)$ , cioè

$$E_{\chi^2}(a_1, \dots, a_m) = \sum_{i=1}^N \left[ \frac{f^{exp}(t_i) - f_{(a_1, \dots, a_m)}^{th}(t_i)}{f_{(a_1, \dots, a_m)}^{th}(t_i)} \right]^2$$

$E_{PDF}$  é diversa dall'energia finale trovata da amoeba (" $E_{END}$ ") in quanto la prima é stata calcolata con  $dt=1$  s, la seconda con  $dt=600$  s. Sono disponibili in letteratura<sup>1</sup> i seguenti valori di  $n_0, \kappa, \lambda, \mu, \alpha, \beta, c^*$  (alcuni sono solo stimati):

---

<sup>1</sup>per  $n_0, \kappa, \lambda, \mu$ : D. Delle Side [7]; per  $\alpha, \beta, c^*$ : P.S. Stewart[8], Trovato et al.[9], Alberghini et al.[10]; si noti che per  $\alpha, \beta, c^*$  i valori riportati sono indicativi in quanto relativi ad altri sistemi di quorum sensing.

$n_0(\frac{\#cellule}{L})$	$\sim 10^8$
$\kappa$	2.70
$\lambda$ (s)	11304
$\mu(s^{-1})$	$1.42 \cdot 10^{-4}$
$\alpha(\frac{molecoleAl}{\#cellule \cdot s})$	$10^2 \div 10^5$
$\beta(h^{-1})$	$\sim 0.07$
$c^*$ (nM)	$\sim 10$

Tabella 6.1: Valori di riferimento di alcuni parametri

Di seguito vengono riportati quattro risultati con i parametri  $\alpha$  e  $A$  ri-convertiti a quelli originari<sup>2</sup> e alcuni grafici esemplificativi:

Risultato #	1	2	3	4
$E_{END}$	0.41	0.42	0.42	0.40
$\kappa$	2.31	1.93	1.66	1.82
$\lambda$ (s)	$2.05 \cdot 10^3$	$2.10 \cdot 10^3$	$1.69 \cdot 10^3$	$3.04 \cdot 10^3$
$\mu(s^{-1})$	$1.41 \cdot 10^{-4}$	$1.16 \cdot 10^{-4}$	$9.60 \cdot 10^{-5}$	$1.17 \cdot 10^{-4}$
$\alpha(\frac{molecoleAl}{\#cellule \cdot s})$	$3.03 \cdot 10^2$	$7.22 \cdot 10^2$	$6.02 \cdot 10^2$	$2.68 \cdot 10^2$
$\beta(h^{-1})$	0.21	0.19	0.29	0.26
$\delta$	0.47	0.78	1.07	0.61
$A(\frac{V}{\#cellule})$	$3.78 \cdot 10^{-6}$	$8.24 \cdot 10^{-6}$	$1.57 \cdot 10^{-5}$	$1.30 \cdot 10^{-5}$
$E_{PDF}$	2.37	0.45	0.44	1.01
$E_{CDF}$	24.54	25.31	25.24	24.88
$E_{\chi^2}$	2.86	1.59	1.29	2.30

Tabella 6.2: Alcuni esempi di risultati

---

<sup>2</sup>Sfruttando le equazioni per  $\tilde{A}$  e  $\tilde{\alpha}$  e i valori di riferimento di  $n_0$  e  $c^*$ .

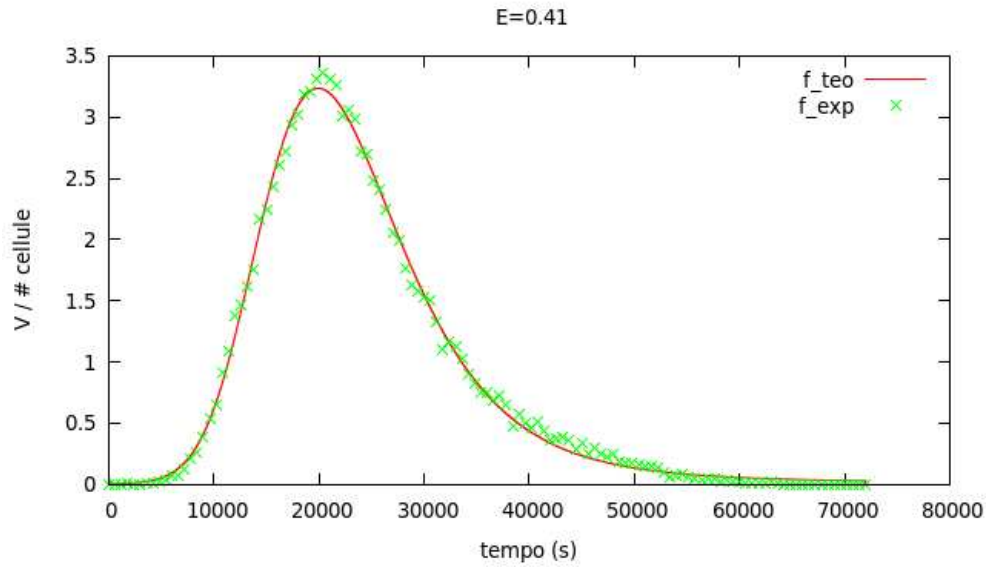


Figura 6.1: Esempio di fit dei dati sperimentali,  $E_{END} = 0.41$

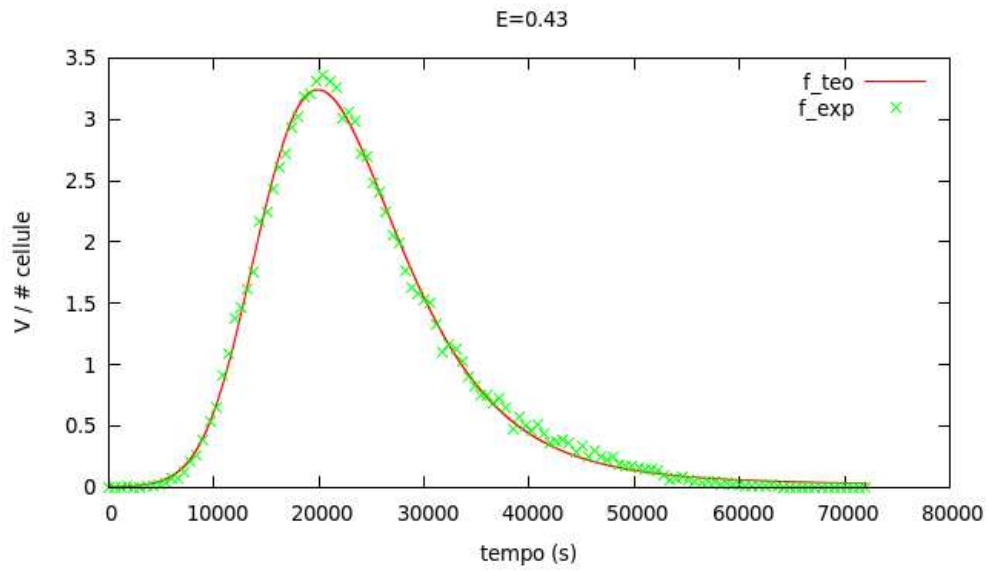


Figura 6.2: Esempio di fit dei dati sperimentali,  $E_{END} = 0.43$

## Capitolo 7

# Conclusioni

I risultati raggiunti, pur essendo corretti nell'ordine di grandezza, sono necessariamente parziali: occorre un cooling rate più alto per esplorare meglio lo spazio dei parametri; questo richiede maggiori risorse computazionali e/o un'ottimizzazione del codice, ad esempio provando con restarts successivi: risulta perciò difficile, al momento, ricavare valori sufficientemente precisi di interesse biologico.

Alternativamente si possono utilizzare algoritmi di ordine superiore al primo in luogo di quello di Eulero qui utilizzato, come il metodo mid-point. Dall'analisi dei risultati si può notare che alcuni parametri rimangono entro un certo ordine di grandezza ( $\alpha$ ,  $\beta, \delta$ ), mentre altri sono più suscettibili a variazioni: per i primi, valori molto distanti dallo zero (soprattutto  $\alpha$  e  $\beta$ ) portano generalmente a soluzioni banali o "infiniti", e questo ha influito nella scelta dei semplici iniziali.

Riguardo ai valori finali della funzione-costo, la differenza fra il risultato  $E_{END}$  dell'ottimizzazione effettuata per  $dt$  alto e il controllo  $E_{PDF}$  calcolato per  $dt$  più basso mostra come  $dt$  vada in linea di principio abbassato per ottenere risultati più affidabili: tuttavia, questa discrepanza non è apprezzabile a livello grafico.

Un'ulteriore incognita è data dal numero di parametri utilizzati: poiché i fit ottenuti sono qualitativamente molto buoni (inoltre il chi-quadro non è mai troppo alto), è lecito chiedersi se i parametri siano sovrabbondanti per la descrizione del problema, cioè se esistano simmetrie interne od opportuni riscalamanti tali da eliminare qualche parametro. Per esempio,  $f(t)$  dipende sia dalla concentrazione di batteri che da quella di AI, dunque potrebbero esservi in generale diverse ripartizioni tali da mantenere inalterato il segnale: sembra che all'aumentare di  $\kappa$ ,  $\delta$  diminuisca. A meno di difficoltà tecniche, ciò suggerisce una risposta affermativa alla domanda precedente.

Il problema della sovrabbondanza dei parametri si potrebbe trattare con un fit simultaneo di più curve sperimentali, mantenendo uguali per tutte le curve quei parametri che si suppone non dipendano dalla particolare colonia

batterica ( $\alpha, \beta, \delta$  e  $A$ , che é lo stesso se le curve provengono da un unico esperimento).

Un'altra questione riguarda il presunto legame tra l'emissione bioluminescente e la distribuzione di Gumbel, menzionata nell'introduzione.

La concentrazione di batteri  $n(t)$  é proporzionale all'esponenziale della funzione di Gompertz: quest'ultima é proprio la CDF della distribuzione di Gumbel. Tenendo conto che  $p(t)$  é il prodotto della funzione di Hill per  $n(t)$  e applicando la regola della derivata del prodotto, il legame può sussistere qualora il contributo della funzione di Hill sia trascurabile.

In conclusione, la comprensione del fenomeno della bioluminescenza rimane ancora un problema aperto e richiede ulteriori contributi modellistici e computazionali: questa tesi mostra le potenzialità e i limiti di una possibile via da seguire.

# Bibliografia

- [1] B.A.Hense M.Schuster. Core principles of bacterial autoinducer systems. *Microbiology and Molecular Biology Reviews*, 2008.
- [2] P.Alifano D. delle Side. Quorum sensing: complexity in the bacterial world. *Chaos, Solitons and Fractals*, 2015.
- [3] D. delle Side M.S.Tredici. Bacterial bioluminescence and gumbel statistics: From quorum sensing to correlation. *Applied Physics Letters*, 2015.
- [4] P.Maire et al. The hill equation: a review of its capabilities in pharmacological modelling. *Fundamental and Clinical Pharmacology*, 2008.
- [5] W.H.Press et al. *Numerical Recipes in C*. Cambridge University Press, second edition, 1992.
- [6] S.Kirkpatrick. Optimization by simulated annealing. *Science*, 1983.
- [7] D.Delle Side. Comunicazione privata.
- [8] P.S. Stewart. Diffusion biofilms. *Journal of Bacteriology*, 2003.
- [9] A.Trovato et al. Quorum vs. diffusion sensing: a quantitative analysis of the relevance of absorbing or reflecting boundaries. *FEMS Microbiology Letters*, 2014.
- [10] S. Alberghini et al. Consequences of relative cellular positioning on quorum sensing and bacterial cell-to- cell communication. *FEMS Microbiology Letters*, 2009.